Random Variables

Mark Schulist

February 21, 2024

Binomial Distribution

We can use `R` to compute probabilities for binomial random variables. Suppose X is a binomial random variable with n trials and probability of success p.

- `dbinom(x, n, p)` computes P(X = x).
- `pbinom(x, n, p)` computes P(X <= x).

In 2021, acceptance rate to WashU was 13% and the graduation rate was 94%.

1. Suppose we randomly sample 20 applicants to WashU. What is the probability that exactly 3 were admitted to WashU?
2. What is the probability that 3 or fewer were admitted to WashU? What is the probability that more than 3 were admitted to WashU?

```r
# P(X = 3)
dbinom(3, 20, 0.13)
```

```
## [1] 0.2347265
```

```r
# P(X <= 3)
pbinom(3, 20, 0.13)
```

```
## [1] 0.7426843
```

```r
# P(X > 3)
1 - pbinom(3, 20, 0.13)
```

```
## [1] 0.2573157
```

```r
pbinom(3, 20, 0.13, lower.tail = FALSE)
```

```
## [1] 0.2573157
```

Exercise 1

If we take a random sample of 100 WashU students, what is the probability that more than 95 of them graduate?

```r
1 - pbinom(95, 100, 0.94)
```

```
## [1] 0.2767752
```

R can also compute factorials and combinations/binomial coefficients.

1. To compute n!, use `factorial(n)`.
2. To compute n choose k, use `choose(n, k)`.

```r
# 6!
factorial(6)
```

```
## [1] 720
```

```r
# 20 choose 3
choose(20, 3)
```

```
## [1] 1140
```

We can also generate random samples from a binomial distribution. `rbinom(x, n, p)` will generate x observations from a binomial random variable with n trials and probability of success p. We can generate a random sample of 20 WashU applicants, and count the number who were admitted:

```r
set.seed(56)
rbinom(1, 20, 0.13)
```

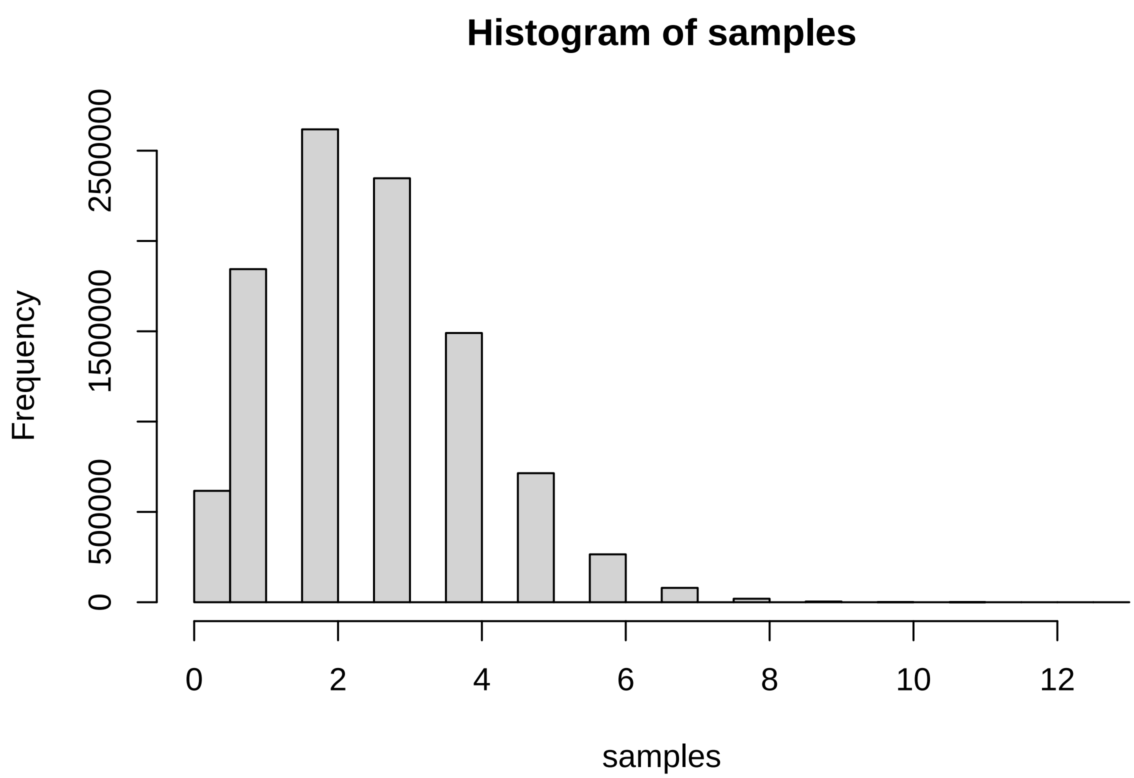
```
## [1] 2
```

We can take repeated random samples by changing the value of x. Each sample provides the number of applicants who were admitted.

```r
samples <- rbinom(10000000, 20, 0.13)
```

Let's make a histogram of `samples` and compute it's mean:

```r
hist(samples)
```




```r
mean(samples)
```

```
## [1] 2.600086
```

The expected number of applicants admitted to WashU in our sample of 20 is $np = 20 \cdot 0.13 = 2.6$

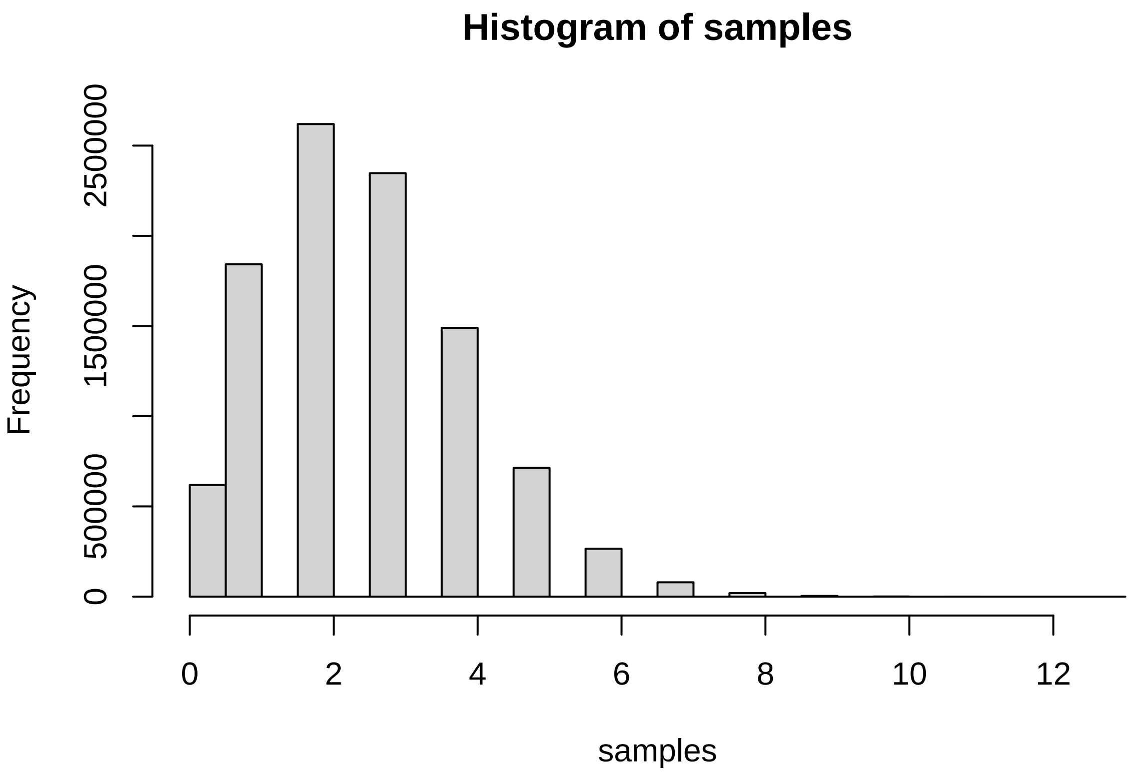
Exercise 2

Take 10000 samples of 20 WashU applicants. Make a histogram of the number of students admitted, and compute the average. How close is the average to 2.6?

```
samples <- rbinom(10000000, 20, 0.13)
mean(samples)
```

```
## [1] 2.599638
```

```
hist(samples)
```



Poisson Distribution

I am interested in investigating the number of students who attend each of my office hours. For 10 office hours, I record the number of students:

1 1 4 2 0 0 2 4 3 1

It might be reasonable to use the Poisson distribution (where we don't know the value of lambda) to model this. Let's start by computing the sample average for the number of students who attended each office hours:

```
students <- c(1, 1, 4, 2, 0, 0, 2, 4, 3, 1)
students
```

```
##  [1] 1 1 4 2 0 0 2 4 3 1
```

```
mean(students)
```

```
## [1] 1.8
```

For a Poisson random variable, the expected value (or average) is lambda. We consider this to be the population average, i.e., the true unknown average number of students who visit each of my office hours. We could use the sample average (1.8) to estimate this unknown population average. We can use the sample average as an estimate of lambda and compute some estimated probabilities.

R has functions to compute probabilities for the Poisson distribution. Suppose X is a Poisson random variable with lambda as the parameter.

- `dpois(x, lambda)` computes $P(X = x)$
- `ppois(x, lambda)` computes $P(X \leq x)$

1. What is the estimated probability that no students come to office hours?
2. What is the estimated probability that 5 or more students come to office hours?

```r
# P(X = 0)
dpois(0, 1.8)
```

```
## [1] 0.1652989
```

```r
# P(X >= 5) = P(X > 4) = 1 - P(X <= 4)
1 - ppois(4, 1.8)
```

```
## [1] 0.03640666
```

```r
ppois(4, 1.8, lower.tail = FALSE)
```

```
## [1] 0.03640666
```

Exercise 3

If I have two chairs for students in my office, what is the probability that some students who come to my office hours will have to stand?

```r
1 - ppois(2, 1.8)
```

```
## [1] 0.2693789
```

```r
ppois(2, 1.8, lower.tail = F)
```

```
## [1] 0.2693789
```